Underwater image and video dehazing with pure haze region segmentation

Simon Emberton, Lars Chittka and Andrea Cavallaro

Centre for Intelligent Sensing, Queen Mary University of London, UK

Abstract

Underwater scenes captured by cameras are plagued with poor contrast and a spectral distortion, which are the result of the scattering and absorptive properties of water. In this paper we present a novel dehazing method that improves visibility in images and videos by detecting and segmenting image regions that contain only water. The colour of these regions, which we refer to as pure haze regions, is similar to the haze that is removed during the dehazing process. Moreover, we propose a semantic white balancing approach for illuminant estimation that uses the dominant colour of the water to address the spectral distortion present in underwater scenes. To validate the results of our method and compare them to those obtained with state-of-the-art approaches, we perform extensive subjective evaluation tests using images captured in a variety of water types and underwater videos captured onboard an underwater vehicle.

Keywords: Dehazing, image processing, segmentation, underwater, white balancing, video processing

1. Introduction

Improving the visibility in underwater images and videos is desirable for underwater robotics, photography/videography and species identification [1, 2, 3]. While underwater conditions are considered by several authors as similar to dense fog on land, unlike fog, underwater illumination is spectrally deprived as water attenuates different wavelengths of light to different degrees [4].

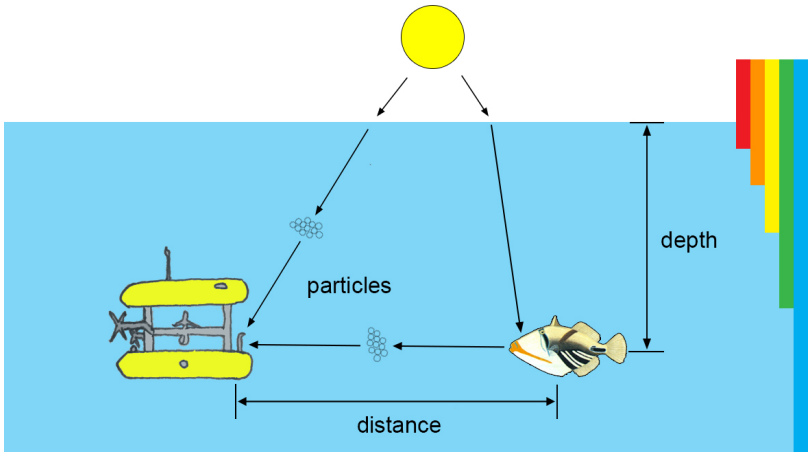


Figure 1: Light is absorbed and scattered as it travels on its path from source, via objects in a scene, to an imaging system onboard an Autonomous Underwater Vehicle.

A key challenge is the spectral distortion in underwater scenes, which dehazing methods are unable to compensate for, especially for scenes captured at depth or in turbid waters (Fig. 1). At depth the distortion is caused by the process of absorption where longer wavelengths (red) are highly attenuated and shorter wavelengths (green and blue) are more readily transmitted [5]. In turbid coastal waters constituents in the water reduce visibility and more readily increase the transmission of green hues [5].

Important parameters to be estimated for underwater dehazing are the veiling light (i.e. the light that is scattered from underwater particles into the line of sight of a camera) and the transmission (i.e. a transfer function that describes the light that is not scattered by the haze and reaches the camera) [6].

Most research has explored dehazing methods for underwater still images [7, 8, 9, 10]. Only few methods have been presented for *dehazing underwater videos* [2, 11, 12]. Applying post-processing to video footage in comparison to images brings new challenges as large variations in estimates between neighbouring frames lead to temporal artefacts. Another challenge for underwater dehazing is to avoid the introduction of noise and false colours. Ancuti *et al.* [2] fuse

contrast-enhanced and grey-world based white-balanced versions of input frames to enhance the appearance of underwater images and videos. This method performs well for most green turbid scenes but the output is often oversaturated and containing false colours, especially in dark regions. Drews-Jr *et al.* [11] use optical flow and structure-from-motion techniques to estimate depth maps, attenuation coefficients and restore video sequences. This method requires the camera to be calibrated in the water which hinders its application to underwater footage captured by other sources [11]. Li *et al.* [12] apply stereo matching to single hazy video sequences and use the generated depth map to aid the dehazing process. Assumptions are made that videos contain translational motions for modelling the stereo and are free of multiple independent moving objects, as only a single rigid motion is considered in the model of 3D structures.

In this work, we propose a method to dehaze underwater images and videos. We address the spectral distortion problem by automatically selecting the most appropriate white balancing method based on the dominant colour of the water. Moreover, to avoid bias towards bright pixels in a scene, we propose to use varying patch sizes for veiling light estimation. We use entropy-based segmention to localise pure haze regions, the presence of which informs the features to select for the veiling light estimate. Finally, to ensure that the pure haze segmentation is stable between video frames, we introduce a Gaussian normalisation step and allocate an image-specific transmission value to pure haze regions to avoid generating artefacts. Then, to ensure coherence across frames, we temporally smooth estimated parameters.

In summary, the main contributions of this work include (i) a semantic white balancing approach for illuminant estimation that uses the dominant colour of the water; (ii) a pure haze detection and segmentation algorithm; (iii) a Gaussian normalisation to ensure temporally stable pure haze segmentation; and (iv) the use of different features generated with varying patch sizes to estimate the veiling light. The main components of the proposed approach are shown in Fig. 2.

The paper is organised as follows. In Section 2 we provide an overview of the state of the art in underwater image dehazing, and white balancing.

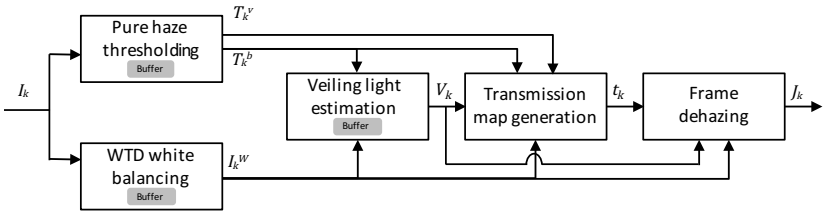


Figure 2: Block diagram of the proposed approach. Temporal smoothing of the parameters takes place in each buffer. KEY – I_k: input frame at time k. T_k^b: Boolean value indicating whether pure haze regions exist; T_k^v: threshold value indicating amount of pure haze in frame; I_k^W: water-type dependent (WTD) white balancing output frame; V_k: veiling light estimate; t_k: transmission map; J_k: dehazed frame.

We introduce methods to white balance underwater scenes (Section 3), select features for the veiling light estimate (Section 4) and segment pure haze regions during the generation of the transmission map (Section 5). In Section 6 we evaluate our approach in comparison to the state of the art in underwater image and video dehazing. Finally, conclusions are drawn in Section 7.

2. State of the art

In this section we discuss the state of the art in image dehazing, the localisation of pure haze regions and finally underwater white balancing methods.

The majority of dehazing approaches make use of the Dark Channel Prior (DCP) [6] or adaptations of it [8, 7, 9]. The DCP makes the assumption that information in the darkest channel of a hazy image can be attributed to the haze [6]. Terrestrial dehazing methods detected sky regions in still images with features such as the saturation/brightness ratio, intensity variance and magnitude of edges [13], gradient information and an energy function optimisation [14], and semantic segmentation [15]. A method for image and video dehazing suppressed the generation of visual artefacts in sky regions by minimising the residual of the gradients between the input image and the dehazed output [16]. An underwater-specific method locates *pure haze regions* by applying a binary adaptive threshold to the green-blue dark channel [17], which fails if the dark

Table 1: Underwater dehazing methods.

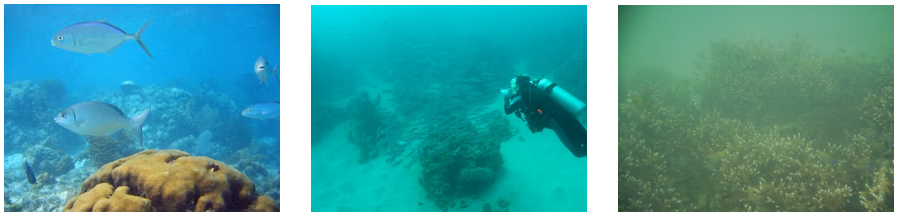
Method	**Dehazing**	**White balancing**	**Applicable to video?**
Chao'10 [18]	Dark Channel Prior [6]	–	–
Chiang'12 [19]	Dark Channel Prior [6]	Fixed attenuation coefficient $K(\lambda)$	–
Ancuti'12 [2]	Fusion	Adapted grey-world	✓
Drews-Jr'13 [7]	Underwater Dark Channel Prior	–	–
Wen'13 [8]	Adapted Dark Channel Prior	–	–
Li'15 [12]	Stereo & Dark Channel Prior	–	✓
Proposed	Adapted Dark Channel Prior	Semantic	✓

channel incorrectly estimates pixels from bright objects as pure haze. An underwater dehazing approach used colour-based segmentation [10], which can be unreliable in cases where, due to the effects of the medium, the colour of objects underwater appear similar to the colour of the pure haze regions.

A number of *underwater-specific methods* [18, 19] estimate the transmission map with the original DCP method [6]. Water attenuates long wavelength (red) light more than short wavelength (blue) light, so the DCP method has the disadvantage that the darkest channel is usually the red channel: when at depth there is no red light and therefore no variation in information in the darkest channel, the transmission estimate will be corrupted [7]. The Underwater Dark Channel Prior [7] solves this limitation by applying DCP only to the green and blue channels.

White-balancing methods include grey-world and white-patch: the grey-world algorithm [20] assumes that the average reflectance in a scene is grey [21]. White-patch [22] makes the assumption that the maximum response in all the channels of a camera is caused by a perfect reflectance (i.e. a white patch) [21]. *Underwater white balancing* methods have assumed that the attenuation coefficient of the water is known [19], which is unlikely in the real world, and have estimated the illuminant using the grey-world assumption and an adaptive parameter that changes depending on the distribution of the colour histogram [2].

Underwater dehazing methods are summarised in Table 1. The main novelties of our proposed approach include stable segmentation of pure haze regions between frames to avoid the production of artefacts in these areas during de-



(a) (b) (c)

Figure 3: Sample images from the PKU-EAQA dataset [23] captured in (a) blue, (b) turquoise and (c) green waters.

hazing, using the presence (or absence) of pure haze to inform the features for the estimation of the veiling light and a semantic approach for white balancing where the method applied is dependent on the water type.

3. Water-type dependent white balancing

We propose a general framework to estimate the dominant colour for the selection of the most appropriate white-balancing method. The proposed method is inspired by semantic methods, which classify images into categories before choosing the most applicable illumination estimation method for each category [21]. We group images into three main classes, namely images captured in blue, turquoise or green-dominated waters (Fig. 3).

Blue-dominated waters are representative of open ocean waters. When white balancing methods (i.e. grey-world, white-patch, max-RGB, grey-edge [24]) are used with images captured in these water types they often introduce undesirable spectral distortions with the colour of the water becoming purple-grey and the colour of objects becoming yellow-grey. There is the potential for white balancing methods to be able to deal with the spectral distortion present in blue waters. However, in this work we avoid white balancing this water type so to inhibit the introduction of spectral distortions.

Turquoise waters contain relatively similar levels of green and blue light than in other spectral domains. For these waters we employ the white-patch

method as the images are likely to fulfil its assumption. Let $I^{WP}(x)$ be a colour-corrected image with the white-patch method and x a pixel location in the image. We employ a white-patch method implemented in CIE L*a*b* space as the lightness channel L^* is intended to approximate the perceptual abilities of the human visual system [25]. We sort the pixel values in order of lightness by transforming the RGB image to CIE L*a*b* space. We denote the estimated illuminant as $E^{WP} = \mu^{Lab}$, where μ^{Lab} is the mean of the pixels in L*a*b* space with respect to the top one percent of pixels in the lightness channel L^* [26].

Green-dominated waters are likely to be found in turbid coastal regions where the waters contain run off from the land and increased levels of phytoplankton and chlorophyll [5]. In these water types light does not travel far and visibility is reduced significantly. The green colour cast is caused by the constituents in the water. These images often still contain useful information in the red channel and it is therefore likely that the grey-world assumption is valid for these cases. Let $I^{GW}(x)$ represent a colour-corrected image with the grey-world method. The grey-world implementation makes use of an illuminant estimation method that is a function of the Minkowski norm [27]. For both illuminant estimation methods, values are converted to CIE XYZ space which describes a colour stimulus in terms of the human visual system [25]. Diagonal transforms are better able to discount the illumination if first transformed to a suitable RGB space [28], therefore we apply a chromatic adaptation transform to white balance the image.

In order to categorise green, turquoise and blue waters, we calculate μ^G and μ^B, the mean value of the green and blue colour channels. Then we consider their difference $\mu^d = \mu^G - \mu^B$. To avoid oversaturated pixels biasing the estimation of the illuminant we apply a median filter to $I(x)$ before white balancing

[21]. We define a white-balanced image as

$$I^W(x) = \begin{cases} \tilde{I}(x) & \text{if } \mu^d \leq -0.05, \\ \tilde{I}^{WP}(x) & \text{if } -0.05 < \mu^d \leq 0.09, \\ \tilde{I}^{GW}(x) & \text{if } \mu^d > 0.09, \end{cases} \quad (1)$$

where our target illuminant is a perfect white light (R=G=B=$\frac{1}{\sqrt{3}}$) [21]. Negative values in Eq. 1 define *blue dominance*, values closer to zero indicate no dominance (i.e. *turquoise waters*) and positive values define *green dominance*. The values of -0.14, -0.01 and 0.13 were calculated for the blue, turquoise and green images in Fig. 3, respectively. We select the thresholds from subjective observations of the performance of the white balancing methods on a training dataset of 44 underwater images. The dataset contained 18 blue (min=-0.259, max=-0.052, mean=-0.129), 9 turquoise (min=-0.048, max=0.015, mean=-0.027) and 17 green images (min=0.110, max=0.601, mean=0.320). Note that we disregard the information in the red channel as it often produces low signals at depth.

4. Veiling light feature selection

The key observation underlying the Dark Channel Prior is that in local patches of an image the minimum intensity can be attributed to haze [6]. Therefore dark channel-based veiling light estimation is biased towards pixels from bright objects in the scene that are larger than the size of the local patch Ω used to generate the dark channel. Large patches help to avoid inaccurate estimates caused by bright pixels which invalidate the prior, while small patches help to maintain image details during the generation of the transmission map [29]. To ensure both global and local information is taken into account we employ a hierarchical veiling light estimation method [10], which fuses a range of layers with different-sized patches. We apply a sliding window approach with varying patch sizes for each layer $l = 1, 2, ..., L$ to generate the veiling light features. If w is the width of the image, we define the size of a local patch at each layer as $\Omega^l = \frac{w}{2^l}$.

We aim to estimate the veiling light from the part in the scene with the densest haze, which is usually found in the most distant location. We use different features depending on the presence of pure haze. We use texture information to avoid bias towards bright objects only in images with pure haze, as it would bias estimates towards textureless regions, such as dark shadows, in images without pure haze.

Compared to areas with objects, pure haze regions have low texture and therefore lower entropy values. Let $G(x)$ be the grey-scale version of $I(x)$. To detect whether an image contains pure haze and then select the features to use in the veiling light estimation accordingly, we compute the entropy image as

$$\eta(x) = 1 - \sum_{y \in \Omega(x)} p(G(y)) \ \log_2 \ p(G(y)), \tag{2}$$

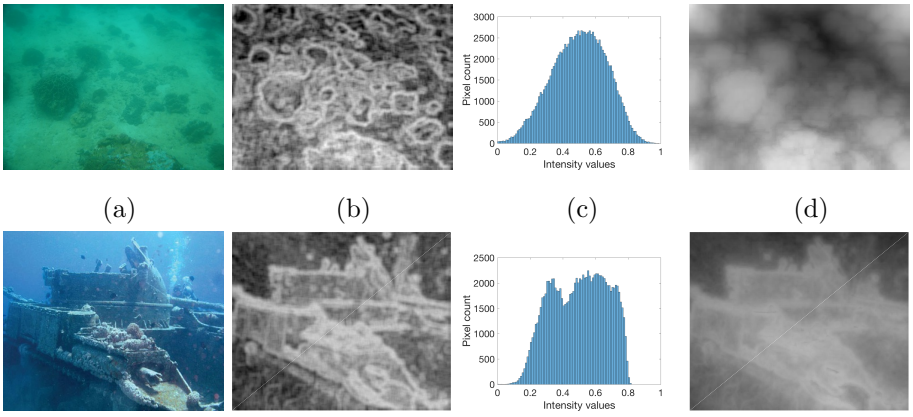
where $p(G(\cdot))$ is the probability of the intensities of G in a local patch $\Omega(x)$ centred at pixel x.

To normalise the values of $\eta(x)$ we propose a method which ensures the mean value of the normalised entropy image $\eta^{'}(x)$ remains stable between consecutive frames. We only use the data within the interval $[\mu - 3\sigma, \mu + 3\sigma]$ and normalise them as

$$\eta^{'}(x) = \begin{cases} 0 & \text{if } \eta(x) < \mu - 3\sigma \\ 1 & \text{if } \eta(x) > \mu + 3\sigma \\ \frac{\eta(x) - (\mu - 3\sigma)}{6\sigma} & \text{otherwise.} \end{cases} \tag{3}$$

Finally, we apply a 2D Gaussian filter (9×9 pixels) to remove local minima and improve spatial coherence, and a 3D Gaussian filter (9×9 pixels and 21 frames) to improve spatial and temporal coherence. The standard deviation of the Gaussian in both filters is set to 1.

We expect the distribution of the histogram of $\eta^{'}(x)$ to be unimodal for scenes without pure haze regions (Fig. 4 (c)) and bimodal for scenes that contain pure haze. In the latter case, the peak containing darker pixels is the pure haze and the peak containing lighter pixels corresponds to foreground objects (Fig. 4 (g)). To determine the difference between the empirical distribution and the unimodal distribution we employ Hartigan's dip statistic [30]. The output of



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4: Pure haze region checker. (a),(e) Sample images and (b),(f) corresponding entropy images. (c),(g) Histograms of the entropy images: a unimodal distribution (top) indicates the absence of pure haze regions; whereas a bimodal distribution (bottom) indicates the presence of pure haze regions. (d) Visualisation of $f(x)$ for veiling light estimation in an image with no pure haze regions. (h) Visualisation of $f^p(x)$ for veiling light estimation in an image with pure haze. Input images taken from the PKU-EAQA dataset [23].

this method is a Boolean threshold value T^b which determines whether an image contains pure haze regions (1) or not (0).

For images *without pure haze regions* we use the green-blue dark channel $\theta^{d^{GB}}(x)$ [7], an adaptation of the Dark Channel Prior [6], which allows us to produce a range map describing the distance of objects in a scene. We generate this feature with varying patch sizes at each layer and then average all the layers together to create our feature image $f(x)$ from which we estimate the veiling light (Fig. 4 (d)).

For images *with pure haze regions* we find the most distant part of the scene to estimate the veiling light using $\theta^{d^{GB}}(x)$ together with the entropy image $\eta(x)$ (Eq. 2). This allows us to avoid bias towards bright pixels. We use $1-\theta^{d^{GB}}(x)$ so that low values indicate pure haze regions for both features. To ensure features are weighted equally we normalise both the features with Eq. 3 before combining them to produce $f^p(x)$ (Fig. 4 (h)).

Finally, we use our feature images, which describe locations likely to contain dense haze, to estimate the veiling light. Let $v = \underset{x}{\text{argmin}}(f(x))$ and Q be the set of pixel positions whose values are v. If Q contains only one element we represent the veiling light V as the RGB value of the white-balanced image $I^W(x)$ in that position. When Q contains multiple pixels we take their mean value in the corresponding positions of $I^W(x)$. Likewise for $v = \underset{x}{\text{argmin}}(f^p(x))$ for images with pure haze regions.

5. Transmission-based pure haze segmentation

A hazy image can be modelled as

$$I(x) = J(x)t(x) + (1 - t(x))V, \tag{4}$$

where $I(x)$ is the observed intensity at x, $J(x)$ is the haze free scene radiance, $t(x)$ is the transmission and V the veiling light [6]. $J(x)t(x)$ describes how scene radiance deteriorates through a medium and $(1 - t(x))V$ models the scattering of global atmospheric light in a hazy environment. Transmission is expressed as $t(x) = \mathrm{e}^{-br(x)}$, where b is the scattering coefficient and $r(\cdot)$ is the estimated distance between objects and the camera [6].

We aim to find the transmission $t(x)$ where the dehazed image $J(x)$ does not suffer from oversaturation due to bright pixels and noise/artefacts in the pure haze regions. We first generate an initial transmission map with the green-blue dark channel $\theta^{d^{GB}}(x)$ [7] (Fig. 5 (a)). Low $t(x)$ values lead to bright pixels in an image becoming oversaturated in the final dehazed image. Therefore, we flag locations where truncation outside of the range [0,1] occurs in either the green or blue colour channels[1] of $J(x)$ [10]. To aid fusion with the initial transmission map we employ a sliding window approach on local patches and for these flagged areas we select the lowest values for $t(x)$ that avoid truncation in $J(x)$.

We treat images with pure haze as a special case in the generation of the transmission map as we only slightly dehaze pure haze regions in order not to

[1] Note that we ignore the red channel as it often contains very low values.

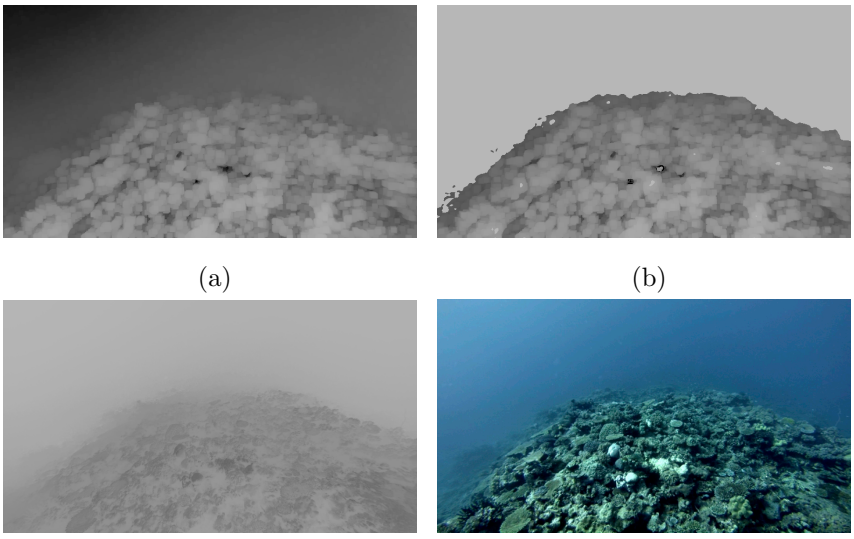


Figure 5: Examples of transmission maps generated from the first frame from sequence S_4 (Fig. 8 (a)). (a) Map estimated with green-blue dark channel [7]. (b) Map with adapted transmission values. (c) Map refined with Laplacian image matting filter [31]. (d) Output frame.

produce artefacts. Pure haze regions can be found through locating the peak of dark pixels in the bimodal distribution of $\eta^{'}(x)$ (Fig. 4 (g)). To automatically find the valley between the two peaks and to ensure only two local maxima remain we iteratively smooth the distribution by thresholding [32]. We define $T^v \in [0,1]$ as the location between these two maxima. For example, for the distribution shown in Fig. 4 (g), $T^v = 0.42$. To ensure small variations within a desirable and limited range $\sim$[0.6,0.7] we allocate pure haze regions a frame-specific transmission value which we define as $\mu^\eta = \mu\left(\eta^{'}(x) > T^v\right)$, and the pure haze-adapted transmission map as

$$t^p(x) = \begin{cases} \mu^\eta & \text{if } \eta^{'}(x) \leq T^v \\ t(x) & \text{if } \eta^{'}(x) > T^v. \end{cases} \tag{5}$$

An adapted transmission map can be seen in Fig. 5 (b) where all the pixels in the pure haze region were assigned a transmission value of 0.66. As the

transmission map requires refinement, we use a Laplacian soft image matting filter [31], which is more suitable to preserve details [29] (Fig. 5 (c)). We use the estimates for the veiling light and transmission map in Eq. 4 to generate the dehazed image (Fig. 5 (d)).

Next, we employ a two-pass moving average filter to smooth T^v and use weights to increase smoothing at neighbouring frames and decrease smoothing for distant frames. We define the smoothed values for the k^{th} frame of T^v, $T^v_\tau(k)$, as

$$T^v_\tau(k) = \frac{1}{(2W+1)^2} \sum_{i=-2W}^{2W} [(2W+1) - |i|]\, T^v(k+i), \tag{6}$$

where W is the number of neighbouring frames either side of $T^v_\tau(k)$ and $2W+1$ is the window size. The value $W = 10$ ensures a good temporal coherence for a frame rate of 25 fps, based on tests on training videos. To maintain temporal coherence for the Boolean threshold value T^b we apply a 21-frame moving average and take the mode of the samples to maintain a binary value.

In addition to T^v we smooth each channel of E^{XYZ} and V with Eq. 6 as variations in these parameters between neighbouring frames (Fig. 6) cause temporal artefacts in the output such as changes in colour and exposure (Fig. 7).

6. Experiments

6.1. Video dehazing analysis

To validate the proposed approach, we compare the proposed method against a fusion method [2] and a method which uses stereo reconstruction to simultaneously dehaze and estimate scene depth [12].

We run the comparison on six diverse and challenging sequences $(S_1, \cdots, S_6)$ of a dataset provided by the Australian Centre for Field Robotics' marine systems group [33]. To reduce processing time and increase the amount of varied footage to be processed by the methods we temporally subsampled all sequences by a factor 4, except for S_2 (subsampled by 8) and S_6 (subsampled by 16). This resulted in videos between 110 and 157 frames. The frames were spatially resized to 768×432 pixels and for this spatial resolution the following parameter

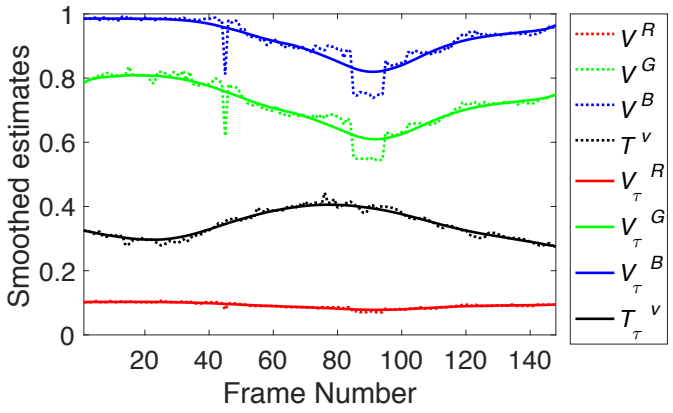


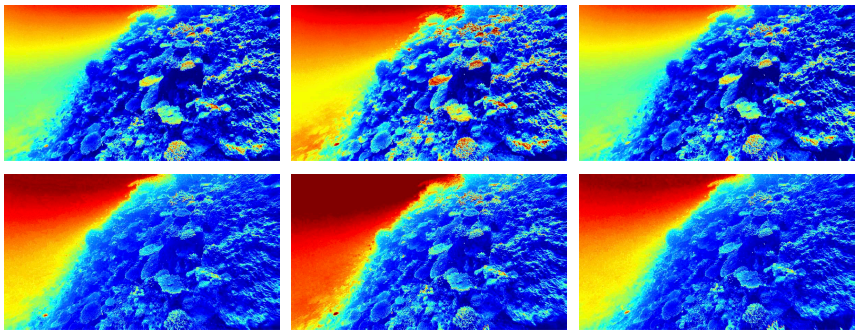
Figure 6: Sample unsmoothed and smoothed estimated parameters. Large variations between neighbouring frames cause temporal artefacts (flicker). This is particularly apparent for the green V^G and blue V^B channels of the veiling light estimates at frames 45, 85 and 95. V^R: red channel of veiling light estimate. T^v: threshold value indicating amount of pure haze. τ: estimate smoothed with Eq. 6.

settings were selected. Before estimating the illuminant the input frame $I(x)$ is spatially smoothed with a median filter with a patch size of 5×5 pixels. A 9×9 patch size is used to generate $\eta(x)$ which is used to locate pure haze regions and $\theta^{dGB}(x)$ which is used as an initial estimate for the transmission map. We use six layers in the veiling light estimation step.

To quantify performance we use UCIQE, Underwater Color Image Quality Evaluation [34], as well as a subjective evaluation. UCIQE is an underwater-specific no-reference metric that takes into account the standard deviation of chroma σ^χ, the contrast of lightness ψ^{L*} and the mean of saturation μ^s. The metric is defined as

$$\text{UCIQE} = \omega_1 \ \sigma^\chi + \omega_2 \ \psi^{L*} + \omega_3 \ \mu^s, \tag{7}$$

where ω_i are the weighted coefficients. Chroma $\chi = \sqrt{a^2 + b^2}$, where a and b are the red/green and yellow/blue channels of the CIE L*a*b* colour space, respectively. Saturation $= \frac{\chi}{L*}$ and ψ^{L*} is defined as the difference between the bottom 1% and the top 1% of pixels in the lightness channel L^* of the CIE L*a*b* colour



(a) (b) (c)

Figure 7: Variation in estimates between neighbouring frames introduces temporal artefacts. Intensity values of the green (row 1) and blue (row 2) colour channels of dehazed frames (a) 44, (b) 45 and (c) 46 of sample sequence with no temporal smoothing (Fig. 6). The veiling light estimate in frame 45 is a different colour and intensity to frames 44 and 46, which leads to a change in colour and exposure in the dehazed frame and is particularly evident in the green and blue channels. Original videos provided by [33].

space. We carried out a subjective evaluation as a single-stimulus subjective test with 12 participants and multiple linear regression to determine the weighted coefficients (ω_1=0.3921, ω_2=0.2853 and ω_3=0.3226) for the individual parts of the measure on the PKU-EAQA dataset of 100 unprocessed underwater images [23, 34].

The *subjective evaluation* of the processed videos consists of a paired-stimulus approach where participants were shown two processed sequences of different methods and asked to decide which they preferred (if any). The experiment was set up as a webpage which meant that it was easier to access a wide range of participants than a traditional laboratory study. It also meant that certain recommendations for subjective evaluation experiments were not followed such as that pertaining to the equipment used for viewing the images (e.g. screen resolution and contrast levels), the environment (e.g. temperature and lighting levels) and the distance to the screen [35]. 36 non-expert participants were shown all of the methods compared with each other for the six video sequences.

Visual inspection of the processed videos reveals that for the method of An-

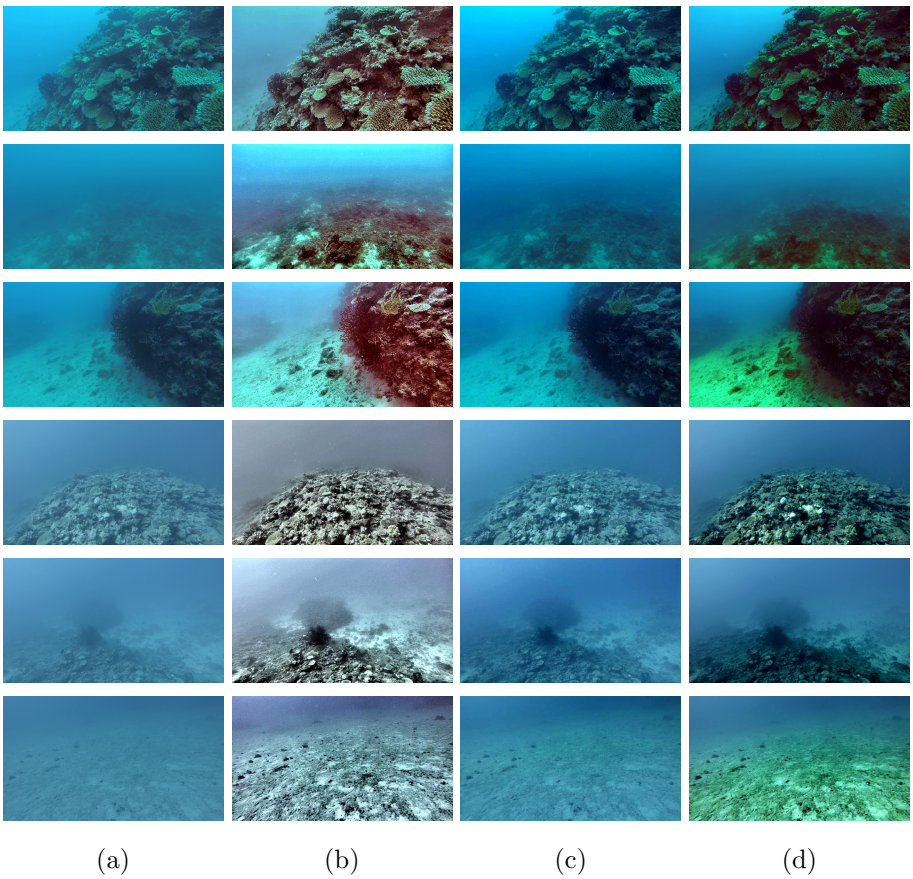
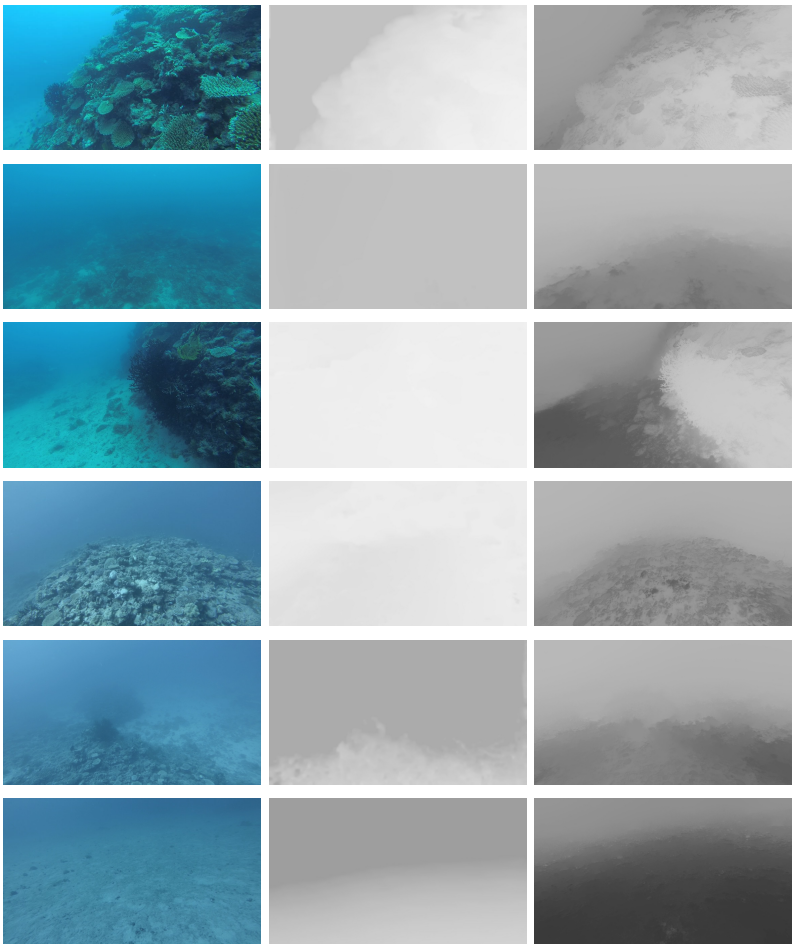


Figure 8: Video results. (a) Input: first frame of sequences S_1-S_6. Output images with (b) [2], (c) [12] and (d) Proposed. Original videos provided by [33].

cuti *et al.* [2] (Fig. 8 (b)) the contrast is over-enhanced giving an oversaturated and unnatural appearance and there are also often red artefacts created in dark regions. The method of Li *et al.* [12] only slightly dehazes the sequences (Fig. 8 (c)), which is evident from the transmission maps (Fig. 9 (b)) that either lack detail or are wrong. The proposed method performs well for most of the sequences (in particular S_4 and S_5) except when they contain large bright objects, such as seabed regions (Fig. 8 (d)). Our method successfully inhibited the production of noise and artefacts in the pure haze regions, unlike the other methods.

It is possible to notice that the main limitation of the proposed method is



(a) (b) (c)

Figure 9: Transmission map results. (a) Input: first frame of sequences S_1-S_6. Transmission maps with (b) [12] and (c) Proposed. Original videos provided by [33].

Table 2: Evaluation results for underwater video methods. UCIQE [34] includes standard deviation of chroma σ^{χ}, contrast of lightness ψ^{L*} and mean of saturation μ^{s}. Subjective evaluation (**SE%**) calculated as percentage preference.

	Ancuti'12 [2]					Li'15 [12]					Proposed				
	σ^{χ}	ψ^{L*}	μ^{s}	UCIQE	**SE%**	σ^{χ}	ψ^{L*}	μ^{s}	UCIQE	**SE%**	σ^{χ}	ψ^{L*}	μ^{s}	UCIQE	**SE%**
S_1	0.44	0.91	0.81	0.69	31.02	0.36	0.71	0.83	0.61	22.22	0.44	0.72	0.87	0.66	46.76
S_2	0.44	0.89	0.76	0.67	37.50	0.33	0.61	0.82	0.57	27.31	0.42	0.61	0.85	0.61	35.19
S_3	0.49	0.88	0.77	0.69	44.44	0.34	0.65	0.85	0.59	23.15	0.40	0.86	0.85	0.68	32.41
S_4	0.23	0.98	0.81	0.63	31.02	0.22	0.57	0.77	0.50	11.57	0.29	0.76	0.86	0.61	57.41
S_5	0.23	0.96	0.74	0.60	39.35	0.18	0.48	0.78	0.46	13.43	0.26	0.64	0.84	0.55	47.22
S_6	0.19	0.94	0.73	0.58	36.57	0.11	0.35	0.76	0.39	20.37	0.25	0.71	0.75	0.54	43.06
Average	0.34	0.93	0.77	0.65	**36.65**	0.27	0.58	0.81	0.53	**19.68**	0.35	0.71	0.85	0.61	**43.67**

with scenes that contain bright objects larger than the patch size, when the transmission map generated with the Dark Channel Prior fails as these nearby objects are incorrectly estimated as being far away. For this reason these objects are given low transmission values and are therefore overly dehazed. Even though our method inhibits the truncation of pixel colour, distortions are created in these areas, e.g. the bright seafloor in S_3 (Fig. 8 (d)). Features such as a depth map created from stereo reconstruction [12] or geometric constraints [36] could be used to improve the transmission map in these situations. Also, no white balancing is applied to these sequences, as they are categorised as blue water type, which may have helped to improve the colour distortion.

Table 2 shows quantitative and subjective evaluation results for the video methods. The proposed approach performs the best overall on the subjective evaluation. However, it is subjectively rated worse than the method of Ancuti *et al.* [2] on S_2 and S_3, because the estimated transmission map introduced spectral distortion in areas containing large bright objects, such as the seafloor.

The proposed method achieves the best scores for μ^{s} and second best on ψ^{L*}. For σ^{χ} the proposed method achieves the best results on S_4, S_5 and S_6. Li *et al.* [12] achieve the worst results on most of the metrics for most of the sequences. The method of Ancuti *et al.* [2] achieves the best overall UQICE score as it attains the best results by far for the contrast of lightness ψ^{L*} on all the sequences. This performance is obtained because of the use of contrast-enhancing filters that ensure a large range between light and dark

Table 3: Quantitative results for the method of Ancuti *et al.* [2] (Fig. 10 (b)) and the proposed method (Fig. 10 (c)) for images 1, 2, 40 and 47 of the PKU-EAQA dataset [23]. Individual parts of UCIQE [34] include standard deviation of chroma σ^{χ}, contrast of lightness ψ^{L*} and mean of saturation μ^{s}.

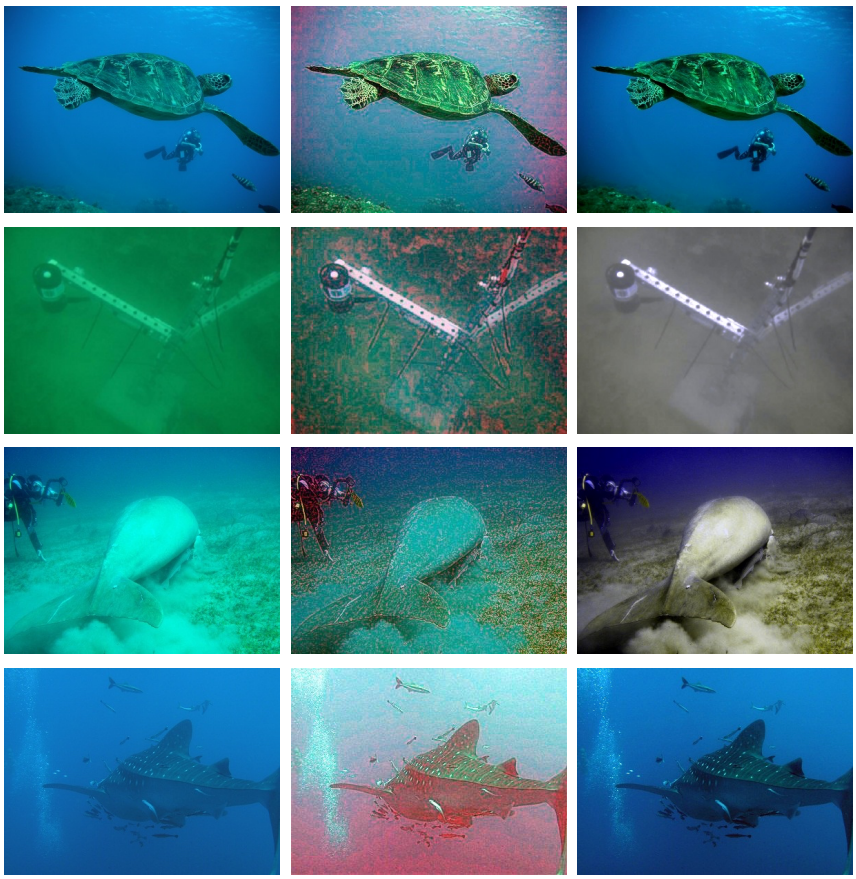
	Ancuti'12 [2]				Proposed			
	σ^{χ}	ψ^{L*}	μ^{s}	UCIQE	σ^{χ}	ψ^{L*}	μ^{s}	UCIQE
Image 1	0.34	0.80	0.78	0.61	0.32	0.65	0.84	0.58
Image 2	0.31	0.67	0.86	0.59	0.13	0.58	0.83	0.48
Image 40	0.32	0.67	0.83	0.58	0.27	0.85	0.89	0.63
Image 47	0.47	0.58	0.73	0.59	0.30	0.50	0.85	0.53

pixels. However, the resulting images become overexposed (e.g. see the grey-white output images in S_4, S_5 and S_6 in Fig. 8 (b)), thus leading to poor results for the mean of saturation μ^{s} and the standard deviation of chroma σ^{χ} for Ancuti *et al.* [2].

Interestingly, the results of σ^{χ} most closely correspond with the subjective evaluation results. However, artefacts and false colours produced by dehazing methods can be counted as positive by the metric. As an example, Table 3 compares UCIQE results for the proposed method (Fig. 10 (c)) and that of Ancuti *et al.* [2] (Fig. 10 (b)) for four images from the PKU-EAQA dataset [23]: the red artefacts produced by the method of Ancuti *et al.* [2] are counted as positive by the σ^{χ} metric and high values of ψ^{L*} lead to a high overall UCIQE score. This analysis demonstrates the importance of conducting a subjective evaluation when assessing the quality of dehazed images and videos.

6.2. Image dehazing analysis

We complement the video evaluation with a large-scale subjective evaluation experiment as well as a comparison with related methods. Our subjective evaluation experiment was completed by a total of 260 participants (both experts and non-experts) on a diverse dataset of underwater images from the PKU-EAQA dataset [23]. Images are captured in a range of water types (62 blue, 22 turquoise and 16 green), depths and contain a variety of objects (humans, fish, corals and shipwrecks). In this subjective evaluation we compare our proposed method to the results of four enhancement methods provided with the dataset,



(a) (b) (c)

Figure 10: Image results provided with the PKU-EAQA dataset [23] where artefacts and false colours can bias quantitative metrics. (a) Input images (1, 2, 40 & 47). Enhanced images with: (b) [2] and (c) Proposed.

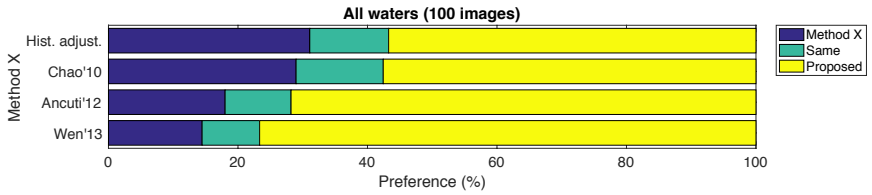


Figure 11: Subjective evaluation results for underwater images. The percentage preference for the proposed method in comparison to other enhancement approaches (Histogram adjustment [23], Chao'10 [18], Ancuti'12 [2] & Wen'13 [8]) for the 100 images in the PKU-EAQA dataset [23]. The χ^2 test results (with a level of significance of 0.05) are 6.33, 13.76, 28.86, and 48.02, respectively. For values > 3.841 the null hypothesis can be rejected. Preference for the proposed method is higher than expected by chance in comparison to all the other methods.

three underwater-specific methods [18, 2, 8] and a histogram adjustment method [23].

We use the same parameter settings and follow the same paired preference experiment detailed in the previous section. As each participant compared random image pairs this resulted in some pairs being evaluated more than others. Each image pair comparing each method was evaluated on average around 26 times (at least 15 times and no more than 36 times). To compare two methods we count the number of times each method is preferred for each image pair with the most popular being awarded a point. No points were awarded when no preference was given for either method.

To detect outliers we followed the β_2 test [35] and none of the participants were rejected. To test the validity our alternative hypothesis (H_a), 'the number of images for which one method performs better than another is higher than expected by chance', we used a Chi-squared test

$$\chi^2 = \sum_{n=1}^{N} \sum_{m=1}^{M} \frac{(O^f_{n,m} - E^f_{n,m})^2}{E^f_{n,m}}, \tag{8}$$

where $O^f_{n,m}$ is the observed frequency at row n and column m; and $E^f_{n,m}$ is the expected frequency for row n and column m. $O^f_{n,m}$ is the total amount of image pairs preferred for each method and $E^f_{n,m}$ is 50 for each method, i.e. those expected by chance. Larger values of χ^2 indicate a stronger evidence for H_a.

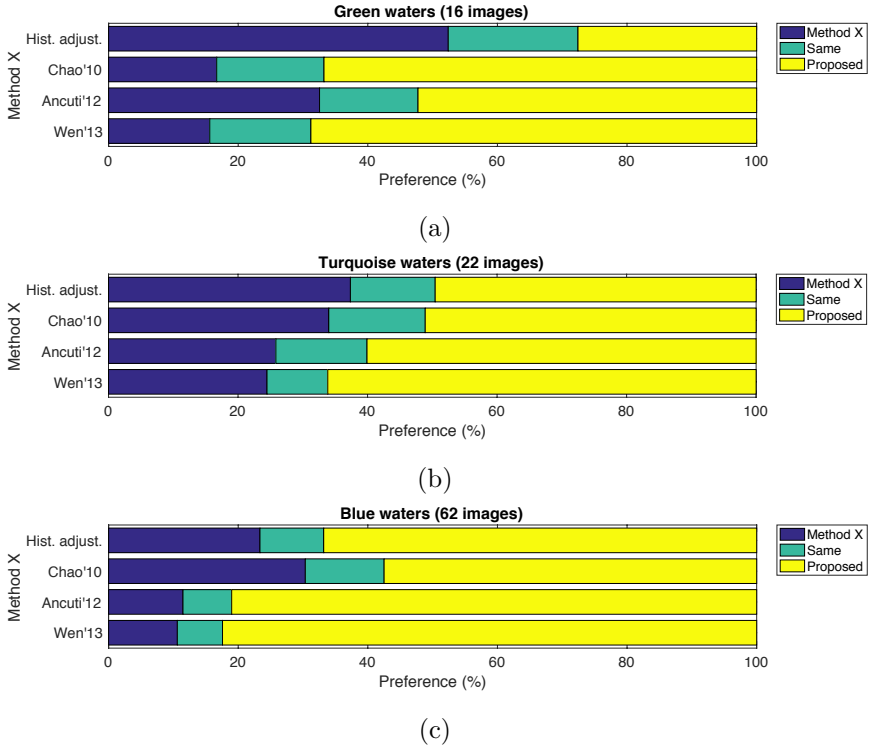


Figure 12: Subjective evaluation results for underwater images arranged into (a) green, (b) turquoise and (c) blue water types. The percentage preference for the proposed method in comparison to other enhancement approaches (Histogram adjustment [23], Chao'10 [18], Ancuti'12 [2] & Wen'13 [8]) for the images in the PKU-EAQA dataset [23].

We set the level of significance to 0.05. If $\chi^2 > 3.841$ the null hypothesis (H_0), 'the number of images for which one method performs better than another is not higher than expected by chance', can be rejected.

The results of our subjective evaluation on enhancement methods for all 100 underwater images (Fig. 11) show there is a preference for the proposed method in comparison to the other methods. This is most pronounced in comparison to the methods of Wen *et al.* [8] and Ancuti *et al.* [2] and less so in comparison to Histogram adjustment and Chao *et al.* [18]. The preference is statistically significant for the proposed method in comparison to the other methods.

Fig. 12 shows that the proposed approach outperforms all the other under-

water dehazing methods for all the waters types. The histogram adjustment method, which was not specifically developed for enhancing underwater scenes, performs the best in green waters. However, this approach performs poorly in blue waters as red artefacts are introduced. These results suggest that a method based on histogram adjustment is promising to address the spectral distortion problem for images captured in green waters.

To complete our analysis, at the end of the subjective evaluation experiment we asked participants to provide details of the selection criteria used when choosing preferred images. Participants were encouraged to tick multiple boxes to indicate the selection criteria they used and 60% chose 'I used different criteria for different images', 47% 'the least blurry', 21% 'the most colourful', 19% 'the most similar to images taken outside water' and 12% 'the brightest'. As we gave participants the option of explaining their selection criteria in more detail, their comments suggest there is a preference for images that are more colourful, higher in contrast and where the visibility of objects in the scene increased as long as the images remained 'natural' and no noise, artefacts and false colours were introduced[2].

We also demonstrate the advantages of our proposed approach through a comparison with two closely-related methods (i.e. the Underwater Dark Channel Prior [7] and our previous method for dehazing single underwater images [10]). The width and height of the images is no larger than 690 pixels. A patch size of 9×9 pixels is used for the generation of the dark channel in all of the methods.

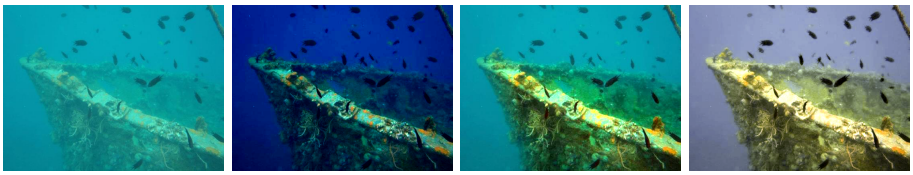
Table 4 shows quantitative results of the three methods with the UCIQE metric [34] previously described in Sec. 6.1 for 10 underwater images previously used to evaluate underwater enhancement methods [10]. Although quantitative measures are not always reliable these results suggest an improved performance for the proposed method.

Fig. 13 compares the dehazed output and Fig. 14 the transmission results for the image 'Galdran9'. The approach of Drews Jr. *et al.* [7] produces dehazed

[2]The results for all of the images and videos can be found in the supplementary material

Table 4: Quantitative results for related methods. UCIQE [34] includes standard deviation of chroma σ^{χ}, contrast of lightness ψ^{L*} and mean of saturation μ^{s}.

	Drews-Jr'13 [7]				Emberton'15 [10]				Proposed			
	σ^{χ}	ψ^{L*}	μ^{s}	UCIQE	σ^{χ}	ψ^{L*}	μ^{s}	UCIQE	σ^{χ}	ψ^{L*}	μ^{s}	UCIQE
Ancuti1	0.23	0.64	0.86	0.55	0.24	0.68	0.84	0.56	0.27	0.81	0.88	0.62
Ancuti2	0.29	0.49	0.88	0.53	0.33	0.40	0.81	0.51	0.35	0.64	0.89	0.61
Ancuti3	0.33	0.64	0.89	0.60	0.38	0.65	0.85	0.61	0.39	0.66	0.88	0.62
Shipwreck	0.24	0.73	0.95	0.61	0.36	0.88	0.87	0.67	0.38	0.99	0.87	0.71
Fish	0.41	0.76	0.91	0.67	0.59	0.85	0.79	0.73	0.60	0.91	0.77	0.74
Galdran1	0.25	0.73	0.91	0.60	0.44	0.90	0.80	0.69	0.38	0.94	0.82	0.68
Galdran9	0.21	0.71	0.95	0.59	0.44	0.85	0.77	0.66	0.37	0.89	0.79	0.65
Reef1	0.32	0.94	0.91	0.69	0.42	0.91	0.84	0.70	0.36	0.97	0.86	0.69
Reef2	0.40	0.84	0.91	0.69	0.61	0.98	0.77	0.77	0.60	0.98	0.76	0.76
Reef3	0.35	0.67	0.92	0.62	0.47	0.95	0.81	0.71	0.43	0.93	0.80	0.69
Average	0.30	0.71	0.91	0.62	0.43	0.80	0.82	0.66	0.41	0.87	0.83	0.68



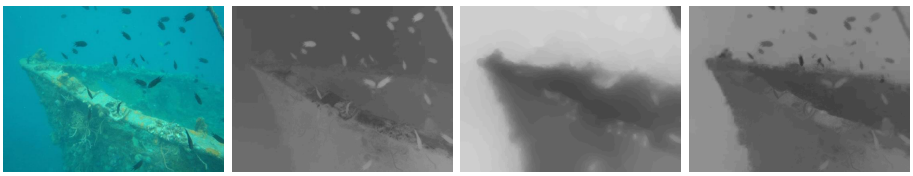
(a) (b) (c) (d)

Figure 13: Image dehazing result. (a) Input image (Galdran9). Output images with (b) [7], (c) [10] and (d) Proposed. Image taken from [9].

output with dark green/blue colour distortions and this is due to the veiling light estimation which is taken from the brightest pixel in the green-blue dark channel (Fig. 13 (b)). In turquoise and green scenes the advantage of the proposed method's white balancing step is demonstrated (Fig. 13 (d)). The transmission refinement method of the proposed approach maintains more image details than the method of Emberton *et al.* [10] and the pure haze segmentation is less prone to failure (Fig. 14).

7. Conclusion

We presented a method for underwater image and video dehazing that avoids the creation of artefacts in pure haze regions by automatically segmenting these areas and giving them an image or frame-specific transmission value. The pres-



(a) (b) (c) (d)

Figure 14: Transmission map result. (a) Input image (Galdran9). Output transmission maps with (b) [7], (c) [10] and (d) Proposed. Image taken from [9].

ence of pure haze is used to determine the features for the estimation of the veiling light. To deal with spectral distortion we introduced a semantic approach which selects the most appropriate white balancing method depending on the dominant colour of the water. Our findings suggest a histogram adjustment method may be advantageous in green-dominated waters. For videos we introduced a Gaussian normalisation step to ensure pure haze segmentation is coherent between neighbouring frames and applied temporal smoothing of estimated parameters to avoid temporal artefacts.

Our approach demonstrated superior performance in comparison to the state of the art in underwater image and video dehazing. Out of three metrics (σ^{χ}, ψ^{L*} and μ^{s} [34]) used to quantitatively assess underwater image quality we found that σ^{χ}, the standard deviation of chroma, most closely corresponded with the subjective evaluation results. However, we also highlighted examples when the results of this metric do not show agreement with subjective assessment as processed images containing false colours are counted as positive by the metric.

The main limitation with our method and, in general, of methods based on the Dark Channel Prior is the incorrect estimation of the transmission map in scenes that contain large bright objects. For this reason our future work will extend our approach to scenes containing large bright objects such as seabed regions and explore white balancing methods that are suitable for all water types.

Acknowledgement

S. Emberton was supported by the UK EPSRC Doctoral Training Centre EP/G03723X/1. Portions of the research in this paper use the PKU-EAQA dataset collected under the sponsorship of the National Natural Science Foundation of China.

References

[1] O. Beijbom, P. J. Edmunds, D. Kline, B. G. Mitchell, D. Kriegman, Automated annotation of coral reef survey images, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1170–1177.

[2] C. Ancuti, C. O. Ancuti, T. Haber, P. Bekaert, Enhancing underwater images and videos by fusion, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 81–88.

[3] M. Roser, M. Dunbabin, A. Geiger, Simultaneous underwater visibility assessment, enhancement and improved stereo, in: IEEE International Conference on Robotics and Automation, 2014, pp. 3840–3847.

[4] P. Emmerson, H. Ross, Variation in colour constancy with visual information in the underwater environment, Acta Psvcholcgka 65 (1987) 101–113.

[5] J. N. Lythgoe, Problems of seeing colours underwater, Plenum Press, New York and London, 1974, Ch. Vision in Fishes: New Approaches in Research, pp. 619–634.

[6] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33(12) (2011) 2341–2353.

[7] P. Drews-Jr, E. Nascimento, F. Moraes, S. Botelho, M. Campos, R. Grande-Brazil, B. Horizonte-Brazil, Transmission estimation in underwater single images, in: IEEE International Conference on Computer Vision Workshop, 2013, pp. 825–830.

[8] H. Wen, T. Yonghong, H. Tiejun, G. Wen, Single underwater image enhancement with a new optical model, in: IEEE International Symposium on Circuits and Systems, 2013, pp. 753–756.

[9] A. Galdran, D. Pardo, A. Picon, A. Alvarez-Gila, Automatic red-channel underwater image restoration, J. Vis. Commun. Image Represent. 26 (2015) 132–145.

[10] S. Emberton, L. Chittka, A. Cavallaro, Hierarchical rank-based veiling light estimation for underwater dehazing, in: British Machine Vision Conference, 2015, pp. 1–12.

[11] P. Drews-Jr, E. R. Nascimento, M. F. M. Campos, A. Elfes, Automatic restoration of underwater monocular sequences of images, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2015, pp. 1058–1064.

[12] Z. Li, P. Tan, R. T. Tan, D. Zou, S. Z. Zhou, L.-F. Cheong, Simultaneous video defogging and stereo reconstruction, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4988–4997.

[13] A. L. Rankin, L. H. Matthies, P. Bellutta, Daytime water detection based on sky reflections, in: IEEE International Conference on Robotics and Automation, 2011, pp. 5329–5336.

[14] Y. Shen, Q. Wang, Sky region detection in a single image for autonomous ground robot navigation, Int. J. Adv. Robot. Syst. 10(362).

[15] K. Wang, E. Dunn, J. Tighe, J. M. Frahm, Combining semantic scene priors and haze removal for single image depth estimation, in: IEEE Winter Conference on Applications of Computer Vision, 2014, pp. 800–807.

[16] C. Chen, M. N. Do, J. Wang, Robust image and video dehazing with visual artifact suppression via gradient residual minimization, in: Proc. of ECCV, 2016, pp. 576–591.

[17] B. Henke, M. Vahl, Z. Zhou, Removing color cast of underwater images through non-constant color constancy hypothesis, in: IEEE International Symposium on Image and Signal Processing and Analysis, 2013, pp. 20–24.

[18] L. Chao, M. Wang, Removal of water scattering, in: IEEE International Conference on Computer Engineering and Technology, Vol. 2, 2010, pp. 2–35.

[19] J. Y. Chiang, Y. C. Chen, Underwater image enhancement by wavelength compensation and dehazing, IEEE Trans. Image Process. 21(4) (2012) 1756–1769.

[20] G. Buchsbaum, A spatial processor model for object colour perception, J. Franklin Inst. 310(1) (1980) 1–26.

[21] A. Gijsenij, T. Gevers, J. Van De Weijer, Computational color constancy: Survey and experiments, IEEE Trans. Image Process. 20(9) (2011) 2475–2489.

[22] E. H. Land, The retinex theory of color vision, Sci. Am. 237(6) (1977) 108–128.

[23] Z. Chen, T. Jiang, Y. Tian, Quality assessment for comparing image enhancement algorithms, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3003–3010.

[24] J. V. D. Weijer, T. Gevers, A. Gijsenij, Edge-based color constancy, IEEE Trans. Image Process. 16(9) (2007) 2207–2214.

[25] G. Wyszecki, W. S. Stiles, Color science (2nd Edition), Wiley, New York, 1982.

[26] G. K. Kloss, Colour constancy using von Kries transformations: Colour constancy 'goes to the lab', Research Letters in the Information and Mathematical Sciences 13 (2009) 19–33.

[27] G. D. Finlayson, E. Trezzi, Shades of gray and colour constancy, in: Color and Imaging Conference, 2004, pp. 37–41.

[28] G. D. Finlayson, S. Susstrunk, Performance of a chromatic adaptation transform based on spectral sharpening, in: Color and Imaging Conference, 2000, pp. 49–55.

[29] S. Lee, S. Y., J.-H. Nam, C. S. Won, S.-W. Jung, A review on dark channel prior based image dehazing algorithms, EURASIP J. Image Video Process. 1 (2016) 1–23.

[30] J. A. Hartigan, P. M. Hartigan., The dip test of unimodality, Ann. Stat. (1985) 70–84.

[31] A. Levin, D. Lischinski, Y. Weiss, A closed-form solution to natural image matting, IEEE Trans. Pattern Anal. Mach. Intell. 30(2) (2008) 228–242.

[32] J. M. Prewitt, M. L. Mendelsohn, The analysis of cell images, Ann. NY Acad. Sci 128(3) (1966) 1035–1053.

[33] ACFR, http://marine.acfr.usyd.edu.au/, last accessed 13th February 2017.

[34] M. Yang, A. Sowmya, An underwater color image quality evaluation metric, IEEE Trans. Image Process. 24(12) (2015) 6062–6071.

[35] ITU-R, Methodology for the subjective assessment of the quality of television pictures, BT.500-13 Edition (2012).

[36] P. Carr, R. Hartley, Improved single image dehazing using geometry, in: IEEE International Conference on Digital Image Computing: Techniques and Applications, 2009, pp. 103–110.